

# Integrated approach for automatic crackle detection based on fractal dimension and box filtering

Cátia Pinho<sup>b,a</sup>; Ana Oliveira<sup>a</sup>; Cristina Jácome<sup>a,c</sup>; João Rodrigues<sup>d</sup>; Alda Marques<sup>a,e</sup>

<sup>a</sup> *Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal.*

<sup>b</sup> *IT – Aveiro, Instituto de Telecomunicações, University of Aveiro, Aveiro, 3810-193, Portugal.*

<sup>c</sup> *Research Centre in Physical Activity, Health and Leisure (CIAFEL), Faculty of Sports, University of Porto, Porto, 4200-450, Portugal.*

<sup>d</sup> *Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Aveiro, Portugal.*

<sup>e</sup> *Institute for Research in Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal.*

## ABSTRACT

Crackles are adventitious respiratory sounds (RS) that provide valuable information on different respiratory conditions. Nevertheless, crackles automatic detection in RS is challenging, mainly when collected in clinical settings. This study aimed to develop an algorithm for automatic crackle detection/characterisation and to evaluate its performance and accuracy against a multi-annotator gold standard. The algorithm is based on 4 main procedures: i) recognition of a potential crackle; ii) verification of its validity; iii) characterisation of crackles parameters; and iv) optimisation of the algorithm parameters. Twenty-four RS files acquired in clinical settings were selected from 10 patients with pneumonia and cystic fibrosis. The algorithm performance was assessed by comparing its results with a multi-annotator gold standard agreement. High level of overall performance (F-score=92%) was achieved. The results highlight the potential of the algorithm for automatic crackle detection and characterisation of RS acquired in clinical settings.

**Keywords:** Adventitious respiratory sounds, Crackles, Discontinuous respiratory sounds, Automatic detection, Classification algorithms, Fractal dimension, Box filtering, Multi-annotator gold standard agreement.

## INTRODUCTION

Acute and chronic respiratory diseases are a major health, societal and economic burden worldwide (World Health Organization, 2008). Therefore, during the last decade, several research efforts have been dedicated to improve the diagnosis, management and monitoring of patients with respiratory diseases (Dinis, Campos, Rodrigues, & Marques, 2012).

Although several measures are available to diagnose and monitor respiratory diseases, conventional auscultation remains the most widely used. The stethoscope has been used for 200 years to perform clinical examinations as it is simple, non-invasive, economic, practical and useful in all populations and settings (Bohadana, Izbicki, & Kraman, 2014). However, it is known that conventional auscultation is subjective, i.e., the interpretation of respiratory sounds depends on the stethoscope properties (Welsby & Earis, 2001), hearing ability and clinical experience of users (Sovijärvi, Vanderschoot, & Earis, 2000) and their capacity to memorise sound patterns (e.g., number and characteristics of adventitious respiratory sounds) (Marques, Bruton, & Barney, 2006).

To overcome the subjectivity associated with conventional auscultation, research efforts have been devoted to improve computerised respiratory sound analysis (CORSa). CORSa consists of recording respiratory sounds with an electronic device and objectively analysing/classifying them based on advanced digital signal processing techniques (Sovijärvi, Vanderschoot, & Earis, 2000). The digitalization of respiratory sounds recordings has the advantages of: i) being compatible with digital signal processing techniques for sound analysis and classification; ii) allowing the precise definition of recording characteristics (e.g., recording level and bandwidth), which can be adjusted to the type of signals that are going to be studied (Cheetham, Charbonneau, Giordano, Helistö, & Vanderschoot, 2000); and iii) overcoming the barriers associated with clinicians having to memorise sound patterns. Through CORSa, respiratory sounds were found to be a more sensitive indicator, detecting and characterising the severity of respiratory diseases before any other measure (Gavriely, Nissan, Cugell, & Rubin, 1994). Furthermore, CORSa can also boost the development of mathematical models of the underlying physical mechanisms of respiratory sound production, to characterise the interaction of mechanical forces, airflow and sound transmission within the respiratory tract (Earis & Cheetham, 2000). This may provide important information to understand and relate different processes of diseases (Earis & Cheetham, 2000).

Computerised respiratory sounds are thus a simple, objective and non-invasive measure to assess the function of the respiratory system (Bohadana et al., 2014). Therefore, special attention has been given to their automatic detection and characterisation, as changes in their properties can early inform the presence of several respiratory conditions (Sovijärvi, A., Malmberg, L., et al., 2000), e.g., pneumonia (Piirila, 1992), bronchiectasis and cystic fibrosis (Marques, Bruton, & Barney, 2009).

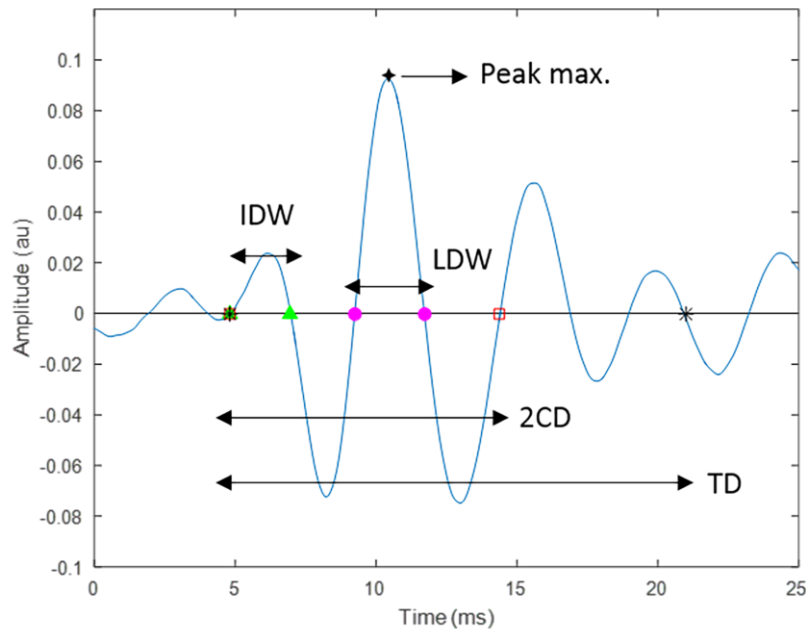
Briefly, respiratory sounds can be classified into normal and adventitious respiratory sounds. Adventitious respiratory sounds are continuous (i.e., wheezes and rhonchus) and/or

discontinuous (i.e., fine crackles and coarse crackles) sounds superimposed on normal respiratory sounds (Bohadana et al., 2014). The specific characteristics of wheezes, such as its duration (longer than 100 ms) and exhibition of distinct peaks in the frequency domain ( $> 100$  Hz) have facilitated the development of methods for its automatic detection and characterisation (Taplidou & Hadjileontiadis, 2007). Crackles, however, are discontinuous, transient and explosive sounds of no more than 20 ms (Sovijärvi, A., Malmberg, L., et al., 2000). These characteristics introduced additional complexity in their automatic detection, and thus further research is needed to overcome this problem.

Crackles occur frequently in cardiorespiratory diseases (Charbonneau et al., 2000; Sovijärvi, A., Malmberg, L., et al., 2000) and their presence has been associated with the severity of diseases. Therefore, the study and detailed analysis of these discontinuous adventitious sounds is important to early detect respiratory diseases and enhance their assessment and monitoring. Previous research has predicted that crackles originating from smaller airways are shorter in duration (i.e., fine crackles) than those originating from larger airways (i.e., coarse crackles) (Sovijärvi, A., Malmberg, L., et al., 2000). Fine crackles are high-pitched inspiratory events that tend to occur in the mid-to-late inspiration and coarse crackles are low-pitched and often occur in the early inspiration and occasionally in expiration as well (Sovijärvi, A., Malmberg, L., et al., 2000). The most studied crackles parameters are:

- the initial deflection width (IDW), which is the duration of the first deflection in a crackle waveform;
- the largest deflection width (LDW), which is the duration of the deflection of the largest amplitude in a crackle waveform;
- the two cycle duration (2CD), which is the time from the beginning of the initial deflection of a crackle to the point where the waveform of the crackle has completed two cycles (Charbonneau et al., 2000; Sovijärvi, A., Malmberg, L., et al., 2000).

These and other parameters are represented schematically in Figure 1.



*Figure 1. Crackle time-domain parameters: initial deflection width (IDW), largest deflection width (LDW), two-cycle duration (2CD), total duration (TD) and maximum peak (peak max.).*

Despite the great value of crackles to contribute for the diagnosis and monitoring of respiratory diseases, their recognition and characterization in respiratory sound files recorded in clinical settings is particularly challenging. Concerning human detection, the level of agreement among health professionals' annotation of crackles has been reported as poor, independently of the health professional background and years of experience (Allingame, Williams, Jenkins, & Tucker, 1995; Brooks & Thomas, 1995; Machado, Oliveira, Aparício, & Marques, 2015). Computerised methods for automatic detection of crackles eliminate this subjectivity (Hadjileontiadis & Rekanos, 2003; Lu & Bahoura, 2008; Vannuccini, Rossi, & Pasquali, 1998). However, the performance of computerised methods is also uncertain, as respiratory sound files acquired in clinical settings have often artefacts (e.g., environmental noise and movement artefacts), which affect the recognition accuracy of algorithms (Chang & Lai, 2010).

Several signal processing techniques have been proposed, including digital filters (Ono, Arakawa, Mori, Sugimoto, & Harashima, 1989), spectrogram analysis (Kaisla et al., 1991), time-domain analysis (Kaisla et al., 1991), auto-regressive models (Hadjileontiadis & Panas, 1996) and wavelet-packet transform methods (Lu & Bahoura, 2008), fuzzy filters (Mastorocostas, Tolia, Theocharis, Hadjileontiadis, & Panas, 2000) and fractal dimension filtering (Hadjileontiadis & Rekanos, 2003). Despite the high values of sensitivity and specificity associated with these techniques, limited testing (Hadjileontiadis & Rekanos, 2003; Lu & Bahoura, 2008; Vannuccini, Rossi, & Pasquali, 1998) have been performed with

respiratory sound files recorded in clinical settings and few have been validated against a multi-annotator gold standard (Quintas, Campos, & Marques, 2013). The gold standard annotation of respiratory sound files can only be achieved by human expert annotation (Marques et al., 2006). Nevertheless, human annotation has an underlying subjectivity (Marques et al., 2006) and therefore, manual annotation performed independently by multiple annotators is essential to produce a more accurate gold standard. Thus, the gold standard must result from the application of statistical agreement criteria to multiple independent annotations obtained for each file (Quintas et al., 2013). Therefore, the acquisition of respiratory sounds in clinical settings and the validation of automatic detection algorithms against a multi-annotator gold standard, are key features to address in the development of more robust algorithms.

Taking these factors into consideration, it may be hypothesised that, when tested in such conditions (i.e., sounds recorded in clinical settings and validation of results against a multi-annotator gold standard), the signal processing techniques proposed in the literature will have lower performances than originally reported. Therefore, a robust algorithm to detect and characterise crackles in respiratory sound files acquired in clinical settings is needed.

This study aimed to develop a new integrated approach for automatic crackle detection and characterisation (based on fractal dimension and box filtering techniques), and to evaluate its performance and accuracy against a multi-annotator gold standard (obtained by the agreement among three experts). Accordingly, this paper is organised in the following four sections: i) *Methods*, which describes the methodologies underlying the development and validation of the proposed algorithm; ii) *Results*, which presents the results of the algorithm's validation in terms of sensitivity (SE); positive predictive value (PPV) and F-score; iii) *Discussion*, which discusses the results and highlights possible improvements; and iv) *Conclusion and future work*, which concludes the paper and presents future work.

## METHODS

The development and evaluation of the algorithm for automatic crackle detection and characterisation was performed using a sample of twenty-four respiratory sound files. Fifteen respiratory sound files were acquired in a community clinical setting from 6 adult patients with pneumonia (3 female; mean age  $46 \pm 14.6$  years old) and 9 respiratory sound files were acquired in a hospital environment from 4 patients with cystic fibrosis (6 female; mean age  $32.3 \pm 18.1$  years old). The selection of these respiratory sound files was based in previous research showing that pneumonia and cystic fibrosis are respiratory diseases characterised by the presence of crackles (Marques et al., 2009; Piirila, 1992). These respiratory sound files were obtained from 2 academic repositories managed by one of the authors of this paper (Alda Marques). Respiratory sounds of patients with cystic fibrosis were collected during a PhD project at the School of Health Sciences of the University of Southampton (SFRH/BD/21375/2005). Respiratory sound files of patients with pneumonia belong to a repository built at the School of Health Sciences of

the University of Aveiro in the scope of a research project (PTDC/SAU–BEB/101943/2008). Approvals for these studies were obtained from institutional ethic committees and respective national data protection.

All respiratory sounds were acquired following CORSA short-term sound acquisition guidelines (Rossi, Sovijärvi, Piirila, Vannuccini, & Dalmasso, 2000). After 5 min of quiet sitting, patients were instructed to breathe normally through the mouth while respiratory sounds were recorded. Recordings were acquired three times in seven chest locations (trachea; left and right: anterior, lateral and posterior), for 25 seconds (Rossi et al., 2000). Two digital stethoscopes have been used: Thinklabs® digital stethoscope (Thinklabs® Rhythm: ds32a, Colorado, US) in patients with pneumonia, and with a WelchAllyn digital stethoscope (WelchAllyn Meditron, 5079-402) in patients with cystic fibrosis. Recordings were performed at a sampling rate (fs) of 44.1 kHz and saved in .wav format. Written informed consents were obtained before any data collection.

## **CRACKLES MANUAL MULTI-ANNOTATION**

Firstly, each file duration was reduced to 10 seconds, because manual annotation is a time-consuming process, since there can be many crackles in a file of few seconds long (Dinis et al., 2012; Oliveira, A., Pinho, Dinis, Oliveira, & Marques, 2013). Then, all respiratory sound files were manually annotated by three respiratory researchers, with experience in visual-auditory crackle recognition, who independently identified the presence of crackles in each respiratory sound file. When the researcher considered that a crackle was present, he/she was requested to annotate the beginning and the end of the crackle. Respiratory Sound Annotation Software V1.1 was used to perform the manual annotation (Dinis et al., 2012).

For each respiratory sound file, a gold standard annotation was obtained by combining the annotations from the three researchers. An event was flagged as a crackle if at least two researchers had identified it, i.e., agreement by majority (Gross, Hadjileontiadis, Penzel, Koehler, & Vogelmeier, 2003). The agreement among the three researchers was set considering the maximum absolute peak within each crackle annotation (Quintas et al., 2013) and was assessed by dividing the number of crackles in which researchers agreed by the total number of crackles annotated. An agreement of 86% was achieved among the three researchers.

## **AUTOMATIC CRACKLE DETECTION**

The proposed algorithm can be summarised in the following main steps:

- i) Extraction of a window of interest of a potential crackle based on fractal dimension estimation (Hadjileontiadis & Rekanos, 2003; Katz, 1988; Lu & Bahoura, 2008) and box filtering techniques (Shen, Shen, Castan, & Zhang, 2002; Taplidou & Hadjileontiadis, 2007);
- ii) Verification of the validity of the potential crackle, considering CORSA established

criteria (Charbonneau et al., 2000; Murphy, Del Bono, & Davidson, 1989);

- iii) Characterisation and extraction of crackle parameters;
- iv) Optimisation of algorithm's parameters.

To reduce computation requirements, the signal was down sampled to 11025 Hz (fs), after appropriate anti-aliasing filtering (i.e., a low-pass digital filter applied to significantly attenuate frequencies above 5512.5 Hz, which would otherwise interfere with the lower frequencies after down sampling). The signal filtering consisted in the application of a passband filter of [100 – 2000] Hz (finite impulse response, designed with a 83ms Blackman window), to eliminate high frequency noise, ensuring that the main features of the crackles were still preserved (Sovijärvi, A., Malmberg, L., et al., 2000; Vannuccini et al., 2000). A detailed description of each one of these steps is given below.

### Extraction of a window of interest

The extraction of a window of interest identifying a possible crackle involved four different signal processing steps:

- i) Savitzky-Golay (polynomial) finite impulse response (FIR) smoothing (Vannuccini et al., 1998);
- ii) Fractal dimension estimation (Hadjileontiadis & Rekanos, 2003; Katz, 1988; Lu & Bahoura, 2008);
- iii) Box filtering (Shen et al., 2002; Taplidou & Hadjileontiadis, 2007);
- iv) Application of a threshold to extract the beginning and the end of a window of interest.

A diagram summarising these steps is presented in

Figure 2.

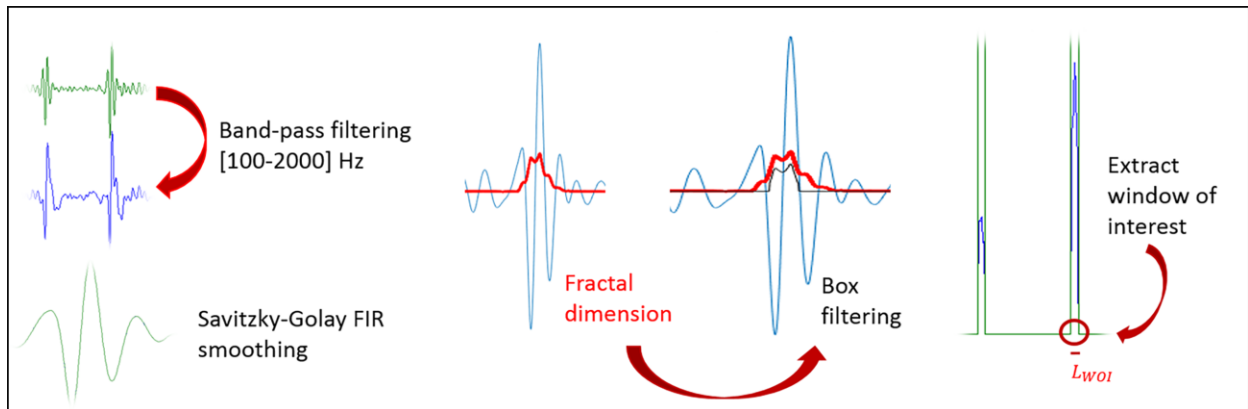


Figure 2. Main steps performed by the algorithm to extract a window of interest of length ( $L_{woi}$ ),

*i.e., identifying a potential crackle.*

Firstly, a smoothing in the signal was performed to extract the remaining high frequency peaks of noise.

Secondly, the fractal dimension of the signal was estimated using a sliding window of length  $L_{FD}$ . Fractal dimension is a statistical measure that indicates an object's complexity in fractal geometry and it is used to evaluate the complexity of a waveform. The method proposed by Sevcik (2010) was used, as it is more sensitive to waveform changes and faster than other commonly used methods, such as the signals envelop method. Fractal dimension is defined by:

$$FD = 1 + \frac{\ln(L)}{\ln(2N')} \quad (1)$$

where  $N'$  is the number of steps in the waveform ( $N' = L_{FD} \times fs - 1$ ) and  $L$  represents the total length of the waveform, i.e., the sum of the Euclidean distance between successive data points:

$$L = \sum_{i=1}^{N'} dist(i, i + 1) \quad (2)$$

Thirdly, an additional smoothing to estimate the trend of the fractal dimension signal in each time instance was applied. The estimation of the trend was based on box filtering, also known as average or mean filtering, with a sliding window of length  $L_{BF}$ . This is a commonly used technique to reduce noise, accomplished by replacing each amplitude sample with the average value of the surrounding samples, including itself. Therefore, the large amplitude variations between samples were removed and a much smoother signal was produced (i.e., the trend). This enhances the identification of crackles in noisy signals, such as sounds acquired in a clinical settings or sounds containing other adventitious respiratory sounds, e.g., wheezes.

In the fourth step, the fractal dimension signal was compared to an adaptive threshold to identify the window of interest of a potential crackle by  $FD > X \times trend$ , where  $X$  is a free parameter for optimisation (Figure 3).



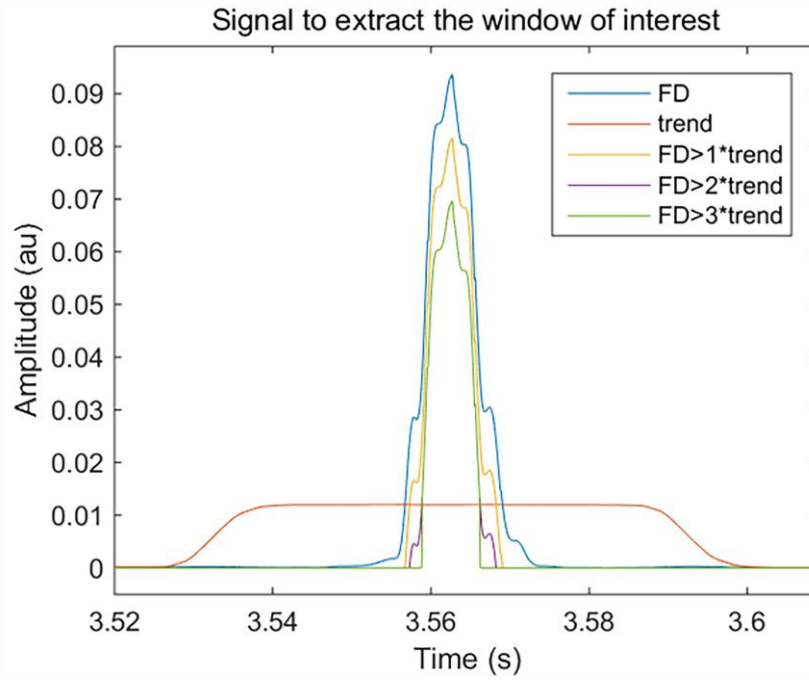


Figure 3. Adaptive threshold applied to the fractal dimension (FD) signal to identify the window of interest of a potential crackle by  $FD > X \times trend$ , where  $X = 1, 2, 3$ .

The final length of the window of interest was defined as the variable  $L_{woi}$ .

### Verification of crackle

A function was developed to evaluate the validity of the potential crackle. This function intended to verify if the window of interest extracted met the CORSA established criteria for crackles detection (Kaisla et al., 1991). Also, additional considerations empirically established by the analysis of noisy signals were implemented to correct errors in the detection. Therefore, the following set of conditions were established:

- i) The amplitude of the different peaks of the crackles had to be progressively lower than the LDW;
- ii) Peaks had to be progressively wider after the IDW (variable defined as  $W_{peaks}$ );
- iii) Crackle zero-crossings were verified: minimum of 5 zero-crossings, to guarantee the calculation of 2CD, and maximum of 16 zero-crossings;
- iv) The mean absolute amplitude of the crackle had to be higher than F times the mean absolute amplitude of the background noise estimated from a segment of length  $L_{bckg}$  preceding the crackle (considering the CORSA criteria, that suggests that crackle amplitude have to be higher than 2\*background noise)
- v) Crackle IDW had to be higher than 1/8 of the LDW.

## Characterisation and extraction of crackle parameters

The following parameters were extracted to characterise the identified crackle: IDW, LDW and the 2CD (Charbonneau et al., 2000; Sovijärvi, A., Malmberg, L., et al., 2000). A detailed description of these crackle time-domain parameters is presented in the Figure 1 (please see the section *Introduction*). Furthermore, the classification of the crackle as fine or coarse was established, based on the criteria proposed by CORSA, i.e., a fine crackle was defined as having a  $2CD \leq 10$  ms, and a coarse crackle has having a  $2CD > 10$  ms (Sovijärvi, A., Dalmasso, F., et al., 2000).

## Optimisation of algorithm's parameters

Seven parameters were chosen to allow tuning the automatic crackle detection algorithm to improve performance. Three different values were tested for each parameter, resulting in a total of 2187 ( $3^7$ ) combinations. The best combination was defined as the one achieving the highest F-score similarity with the gold standard annotation agreement.

The chosen tuning parameters and corresponding tested values were:

- 1) The width of the fractal window of approximately half the length of the crackle (Lu & Bahoura, 2008) ( $L_{FD} = 4, 6, 8$  ms);
- 2) The width of the box filtering window ( $L_{BF} = 40, 60, 80$  ms);
- 3) The  $X$  parameter applied in the adaptive thresholding step to identify the window of interest ( $X = 1, 2, 3$ );
- 4) The minimal length of the window of interest of each crackle ( $L_{WOI} = 0.5, 1, 2$  ms);
- 5) The widening tolerance range ( $= 1/5, 1/4, 1/3$ ) associated with the width of the peaks following IDW ( $W_{peaks} \pm \text{widening tolerance} \times W_{peaks}$ );
- 6) The length of the background sound before the crackle ( $L_{bckg} = 20, 40, \text{all}^*$  ms)  
\*considering all background length before the crackle;
- 7) The factor multiplied by the amplitude of the background sound ( $F = 1.6, 1.8, 2$ ).

Parameters 6 and 7 were established to evaluate the amplitude of the crackle when compared with the background noise (condition 'iv' defined in the section *Verification of crackle*). Only the background length before the crackle window was evaluated. This strategy was adopted to allow the extraction of crackles previously detected in that signal, ensuring that only the fluctuations of the background sound were considered and not the superposition of adventitious sounds (i.e., crackles).

Figure 4 provides the flow diagram of the proposed integrated approach to automatically detect crackles.

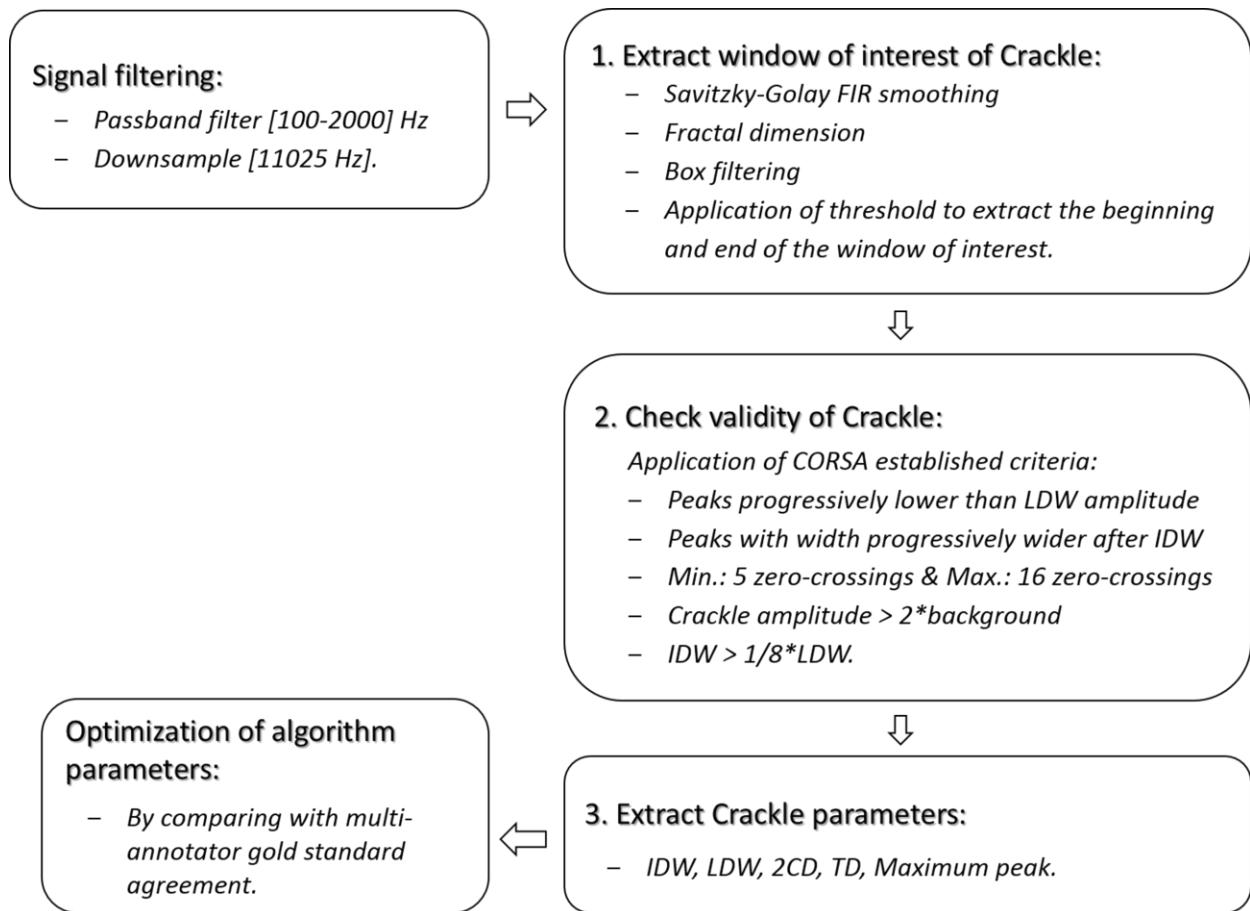


Figure 4. Flow diagram of the proposed algorithm to detect crackles.

The algorithm regarding the proposed integrated approach was written in Matlab®R2009a (Mathworks, Natick, MA, USA).

## CLASSIFICATION MEASURES TO VALIDATE THE ALGORITHM

The validation of the performance and accuracy of the automatic crackle detection algorithm was obtained by comparing the maximum absolute peak within each identified crackle with the results of the multi-annotator gold standard agreement, which for respiratory sounds is the judgment of respiratory experts (please see section *Crackles manual annotation*).

Formally, the proposed classification system is based in two classes (Cheetham et al., 2000), i.e., each event is mapped with a positive (P) (i.e., a crackle) or negative (N) (i.e., not a crackle or non-crackle) class label. The classification model then maps events to predicted classes (Fawcett, 2004, 2006). The event measure is given by the multi-annotator gold standard results, and the classifier refers to the algorithm outcomes. Thus, given a classifier and an event, it is possible to count four well-known parameters:

- i) True positive (TP): if the event is flagged as crackle (i.e., positive) by the gold standard and classified as positive by the algorithm;
- ii) False negative (FN): if the event is flagged as positive by the gold standard and classified as negative (i.e., non-crackle) by the algorithm.
- iii) True negative (TN): if the event is flagged as negative by the gold standard and classified as negative by the algorithm;
- iv) False positive (FP): if the event is flagged as negative by the gold standard and classified as positive by the algorithm;

A two-by-two confusion matrix is used to describe this classification system and provide information regarding the actual and predicted classifications (Akay, 2009; Fawcett, 2006), Table 1.

*Table 1. Confusion matrix representation.*

<b>Actual</b>	<b>Predicted</b>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	True positive ( <b>TP</b> )	False negative ( <b>FN</b> )
<i>Negative</i>	False positive ( <b>FP</b> )	True negative ( <b>TN</b> )

This matrix forms the basis for many common metrics. Sensitivity (SE) and specificity (SP) are two measures that estimate a classifier's performance on different classes (Fawcett, 2006; Sokolova, Japkowicz, & Szpakowicz, 2006).

SE is defined as the ratio between crackles correctly detected (TP) and the sum of TP with the number of crackles not detected by the algorithm (FN), eq. (3).

$$SE = \frac{TP}{TP + FN} \quad (3)$$

SP is defined as the ratio between the events correctly classified as not a crackle (TN) and the sum of TN with the number of crackles incorrectly detected by the algorithm (FP), eq. (4).

$$SP = \frac{TN}{TN + FP} \quad (4)$$

As the analysed files are composed by 110250 points (i.e.,  $10s \times 11025s^{-1}$ ), even for a maximum number of crackles (e.g., 200), the order of magnitude of TN will be around 500 times above the FP count, and thus the SP value will always be approximately 100%. Thus, SP metrics were not

considered in the results.

Other classification metrics is the positive predictive value (PPV) (Akay, 2009; Fawcett, 2006). PPV is the ratio between TP and the total number of crackles detected by the algorithm, i.e., correctly detected (TP) and incorrectly detected (FP), eq. (5).

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

The total performance F-score (F) is a multi-class classification metrics that combines PPV and SE, presenting, therefore, a more robust measure of the algorithm's performance, see eq. (6).

$$F = \frac{(1 + \beta^2) \times SE \times PPV}{(\beta^2 \times PPV) + SE} \quad (6)$$

The parameter  $\beta$  allows to weight more heavily SE or PPV, i.e.,  $\beta > 1$  favours PPV and  $\beta < 1$  SE (Hripcsak & Rothschild, 2005; Sokolova et al., 2006). In the F-score defined in the proposed classification system was assumed  $\beta = 1$ , i.e., an evenly balance between SE and PPV (Hripcsak & Rothschild, 2005; Sokolova et al., 2006), eq. (7).

$$F = 2 \times \frac{SE \times PPV}{SE + PPV} \quad (7)$$

## RESULTS

From a total of 2187 combinations of the threshold parameters, the performance of the automatic crackle detection algorithm was optimised with the following set of parameters:  $L_{FD} = 6 \text{ ms}$ ;  $L_{BF} = 60 \text{ ms}$ ;  $X = 3$ ;  $L_{WOI} = 2 \text{ ms}$ ; *widening tolerance* =  $1/4$ ;  $L_{bckg} = 20 \text{ ms}$  and  $F = 2$ . A summary of the results obtained with this set of parameters is presented in Table 2.

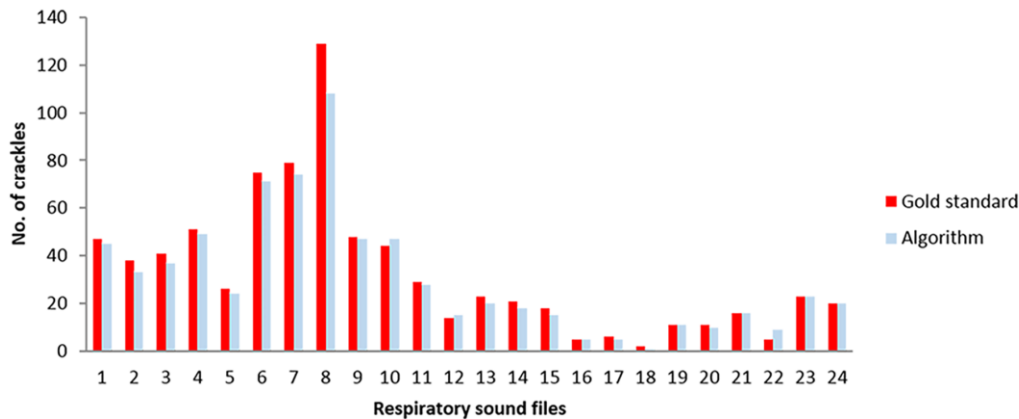
*Table 2. Results obtained for the 24 respiratory sound files analysed.*

N. File	Algorithm	GS	TP	FP	FN	SE	PPV	F
1	45	47	44	1	3	0.94	0.98	<b>0.96</b>
2	33	38	33	0	5	0.87	1.00	<b>0.93</b>
3	37	41	36	1	5	0.88	0.97	<b>0.92</b>
4	49	51	48	1	3	0.94	0.98	<b>0.96</b>
5	24	26	24	0	2	0.92	1.00	<b>0.96</b>
6	71	75	69	2	6	0.92	0.97	<b>0.95</b>
7	74	79	73	1	6	0.92	0.99	<b>0.95</b>
8	108	129	108	0	21	0.84	1.00	<b>0.91</b>
9	47	48	46	1	2	0.96	0.98	<b>0.97</b>

10	47	44	43	4	1	0.98	0.91	<b>0.95</b>
11	28	29	27	1	2	0.93	0.96	<b>0.95</b>
12	15	14	14	1	0	1.00	0.93	<b>0.97</b>
13	20	23	20	0	3	0.87	1.00	<b>0.93</b>
14	18	21	18	0	3	0.86	1.00	<b>0.92</b>
15	15	18	15	0	3	0.83	1.00	<b>0.91</b>
16	5	5	5	0	0	1.00	1.00	<b>1.00</b>
17	5	6	5	0	1	0.83	1.00	<b>0.91</b>
18	1	2	1	0	1	0.50	1.00	<b>0.67</b>
19	11	11	11	0	0	1.00	1.00	<b>1.00</b>
20	10	11	9	1	2	0.82	0.90	<b>0.86</b>
21	16	16	15	1	1	0.94	0.94	<b>0.94</b>
22	9	5	4	5	1	0.80	0.44	<b>0.57</b>
23	23	23	23	0	0	1.00	1.00	<b>1.00</b>
24	20	20	18	2	2	0.90	0.90	<b>0.90</b>

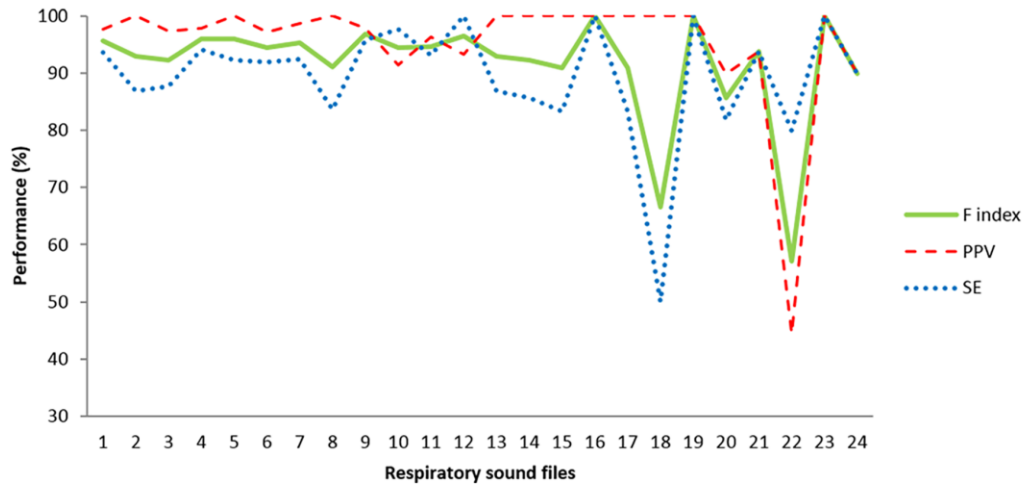
***N. File:** number of sound files analysed; **Algorithm:** number of crackles detected by the proposed automatic crackle detection algorithm; **GS** – gold standard: counts of crackles identified by the multi-annotator gold standard agreement; **TP** – true positive: counts of crackles correctly detected; **FP** – false positive: counts of crackles incorrectly detected; **FN** – false negative: counts of non-crackles incorrectly identified; **SE** – sensitivity, **PPV** – positive predictive value, and **F** – F-score, present the results obtained for the three different classification metrics assessed to validate the algorithm performance.*

The number of crackles identified with the optimised algorithm (from 1 to 108) was similar to the one obtained by the multi-annotator gold standard (from 2 to 129). These results, per respiratory sound file, are shown in Figure 5.



*Figure 5. Crackles detected by manual annotation (multi-annotator gold standard agreement) and by the proposed algorithm for each respiratory sound file.*

Figure 6 presents the SE, PPV and F-score of the algorithm for each respiratory sound file. It can be observed that files 18 and 22, acquired in patients with pneumonia, were those in which the algorithm presented the lowest performance, F-score of 57% and 67% respectively.



*Figure 6. Sensitivity (SE), positive predictive value (PPV) and F-score of the proposed algorithm for each respiratory sound.*

However, when considering the 24 respiratory sound files, the average of SE, PPV and F-score of the proposed algorithm were considerably high ( $SE=89 \pm 10\%$ ,  $PPV=95 \pm 11\%$  and  $F\text{-score}=92 \pm 10\%$ ).

## DISCUSSION

Crackles are adventitious respiratory sounds that frequently occur in cardiorespiratory diseases (Charbonneau et al., 2000; Sovijärvi, A., Malmberg, L., et al., 2000). Thus, their automatic detection and detailed analysis is of major importance for the diagnosis and monitoring of respiratory diseases. Nevertheless, the techniques provided in the literature for this calculus, e.g., (Hadjileontiadis & Rekanos, 2003; Lu & Bahoura, 2008; Vannuccini et al., 1998), are poorly tested in clinical settings and not validated against a multi-annotator gold standard agreement.

Therefore, an integrated approach for automatic crackle detection has been proposed in this study to overcome the identified problems, based on:

- i) The use of the combination of fractal dimension estimation and box filtering techniques to enhance the recognition of crackles in noisy signals;
- ii) The establishment of a set of rules to verify the validity of the detected crackle, grounded by CORSA established criteria (Charbonneau et al., 2000; Murphy et al., 1989);
- iii) The use for a multi-annotator gold standard manual annotation to validate the performance

- of the proposed automatic detection algorithm;
- iv) The optimisation of seven parameters established to improve the performance of the automatic crackle detection algorithm.

The results obtained from the optimisation of the seven parameters tested in a sample of twenty-four respiratory sound files acquired in clinical settings, demonstrated that this was an efficient and robust method for crackle detection/classification and highlights its potential to be used in respiratory sound files acquired in clinical practice.

The conditions initially implemented in the algorithm strictly followed the ones proposed by the CORSA criteria. However, during the iterative assessment of the algorithm, a sharp divergence between the results obtained with the algorithm and those from the gold standard was verified. To address this problem, and as CORSA criteria are not updated since 2000, some refinements of the standard rules were implemented. For example, in the criteria stating that the width of the peaks should be progressively wider after IDW, a deviation of 25% was allowed. It should be noted, that despite the high subjectivity associated with human detection of crackles, this is still considered the only valid method for its detection (Marques et al., 2006), as health professionals are the ones who use it routinely to establish diagnosis and monitoring patients. Therefore, from the authors' point of view, algorithms should be developed to always match human gold standard annotation and not the opposite. Though, as human annotation has an underlying subjectivity the gold standard annotation should be provided by the statistical agreement criteria of multiple independent annotations (i.e., three or more) obtained for each sound file. Thus, in this study a multi-annotator gold standard agreement from three independently manual annotation experts was established.

The SE (89%), PPV (95%) and F-score (92%) of the proposed algorithm are comparable or even higher than those of other previously published methods (SE 80–91%; PPV 83–88%; F-score 86.7%) (Quintas et al., 2013; Vannuccini et al., 1998). However, it should be noted that, contrarily to other algorithms, the presented one was tested with respiratory sound files recorded in clinical settings, which further increase its potential to be used by health professionals in their daily practice. Only in two respiratory sound files from patients with pneumonia, the performance of the algorithm did not reach these high standard values (F-score of 57% and 67%). Such findings are related with the low number of crackles presented (TP) in these two respiratory sound files (1–9 crackles in the sound files), causing the few FP and FN to have a negative impact in the F-score. Nevertheless, this finding may not be clinically relevant as it has been reported that even healthy people present approximately 4 crackles per breathing cycle (Oliveira, Ana & Marques, 2014). Hence, errors of this magnitude might not be clinically significant for diagnosis and monitoring of respiratory diseases.



## **LIMITATIONS**

This study has some limitations that need to be acknowledged. The proposed algorithm was developed by comparing the results against a multi-annotator gold standard obtained from the annotations of three respiratory researchers. Despite the high agreement achieved (86%), it is well-known that human annotation is associated with high levels of subjectivity, thus future studies should consider creating a repository of respiratory sounds annotated by an additional number of experts (e.g., five or more) to minimise bias. Also, a small sample of respiratory sounds files was included from adult patients with pneumonia and cystic fibrosis. Therefore, it would be a great asset to validate the proposed algorithm in large sets of data from patients with other respiratory diseases and different age ranges (young children and infants). Finally, although the proposed algorithm is currently processing in real-time, making it suitable to be used in the clinical practice, would benefit from being included in Clinical Decision Support Systems. This would allow comparisons between patterns of samples and provide info/pre-diagnosis to the caregiver (Musen, Middleton, & Greenes, 2014), which is not addressed in the current version of the algorithm.

## **CONCLUSION AND FUTURE WORK**

The proposed integrated approach for automatic crackle detection achieved a high performance (F-score 92%). This promising result highlights the potential of this new approach for automatic crackle detection and classification in respiratory sounds acquired in clinical settings. Future research could also test the developed algorithm in home-based environments and to be used in patients' tele-monitoring devices.

It would also be of great interest to develop a portable technology including a sound acquisition device plus an open source software integrating the developed algorithm. This would allow health professionals to have instant feedback on their respiratory sound recordings. Additionally, international collaborations are being established to acquire a large database of normal and adventitious respiratory sounds (including crackles, wheezes, rhonchi, stridor, etc.) that could be used to find unsuspected relationships between normal/adventitious respiratory sounds and respiratory diseases, summarise respiratory data into patterns and thus be used in Clinical Decision Support Systems.

## **ACKNOWLEDGEMENTS**

This work was funded by Fundação para a Ciência e a Tecnologia, Portugal -Project PTDC/SAU-BEB/101943/2008, by the ERDF -European Regional Development Fund through the COMPETE Program (operational program for competitiveness) (ref. FCOMP-01-0124-FEDER-013049) and the National Strategic Reference Framework (ref. BI/UI64/3902/2010). The authors would also like to acknowledge João Dinis for his inputs in the development of the

algorithm.

## REFERENCES

- Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, 36(2, Part 2), 3240-3247.
- Allingame, S., Williams, T., Jenkins, S., & Tucker, B. (1995). Accuracy and reliability of physiotherapists in the interpretation of tape-recorded lung sounds. *Aust J Physiother*, 41(3), 179-184.
- Bohadana, A., Izbicki, G., & Kraman, S. S. (2014). Fundamentals of lung auscultation. *The New England journal of medicine*, 370(8), 744-751.
- Brooks, D., & Thomas, J. (1995). Interrater reliability of auscultation of breath sounds among physical therapists. *Physical therapy*, 75(12), 1082-1088.
- Chang, G. C., & Lai, Y. F. (2010). Performance evaluation and enhancement of lung sound recognition system in two real noisy environments. *Computer Methods and Programs in Biomedicine*, 97(2), 141-150.
- Charbonneau, G., Ademovic, E., Cheetham, B., Malmberg, L., Vanderschoot, J., & Sovijärvi, A. (2000). Basic techniques for respiratory sound analysis. *European Respiratory Review*, 10(77), 625-635.
- Cheetham, B., Charbonneau, G., Giordano, A., Helistö, P., & Vanderschoot, J. (2000). Digitization of data for respiratory sound recordings. *European Respiratory Review*, 10(77), 621-624.
- Dinis, J., Campos, G., Rodrigues, J., & Marques, A. (2012). *Respiratory Sound Annotation Software*. Paper presented at the International Conference on Health Informatics, Vilamoura, Portugal.
- Earis, J., & Cheetham, B. (2000). Future perspectives for respiratory sound research. *European Respiratory Review*, 10 641-646.
- Fawcett, T. (2004). ROC Graphs: Notes and Practical Considerations for Researchers. In H. Laboratories (Ed.), (pp. MS 1143-1501). Netherlands.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- Gavriely, N., Nissan, M., Cugell, D. W., & Rubin, A. H. (1994). Respiratory health screening using pulmonary function tests and lung sound analysis. *European Respiratory Journal*, 7(1), 35-42.
- Gross, V., Hadjileontiadis, L. J., Penzel, T., Koehler, U., & Vogelmeier, C. (2003). *Multimedia Database "Marburg Respiratory Sounds (MARS)"*. Paper presented at the 25<sup>a</sup> Annual International Conference of the IEEE EMBS, Cancun, Mexico.
- Hadjileontiadis, L. J., & Panas, S. M. (1996, 31 Oct-3 Nov 1996). *Nonlinear separation of crackles and squawks from vesicular sounds using third-order statistics*. Paper presented at the Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE.

- Hadjileontiadis, L. J., & Rekanos, I. T. (2003). Detection of explosive lung and bowel sounds by means of fractal dimension. *Signal Processing Letters, IEEE*, 10(10), 311-314.
- Hripesak, G., & Rothschild, A. S. (2005). Agreement, the F-Measure, and Reliability in Information Retrieval. *Journal of the American Medical Informatics Association : JAMIA*, 12(3), 296-298.
- Kaisla, T., Sovijärvi, A., Piirila, P., Rajala, H. M., Haltsonen, S., & Rosqvist, T. (1991). Validated Method for Automatic Detection of Lung Sound Crackles. *Medical & Biological Engineering & Computing*, 29(5), 517-521.
- Katz, M. J. (1988). Fractals and the analysis of waveforms. *Computers in Biology and Medicine*, 18(3), 145-156.
- Lu, X. G., & Bahoura, M. (2008). An integrated automated system for crackles extraction and classification. *Biomedical Signal Processing and Control*, 3(3), 244-254.
- Machado, A., Oliveira, A., Aparício, J., & Marques, A. (2015). Respiratory auscultation: (dis)agreement between health professionals. *European Respiratory Journal*, 46(suppl 59).
- Marques, A., Bruton, A., & Barney, A. (2006). Clinically useful outcome measures for physiotherapy airway clearance techniques: a review. *Physical Therapy Reviews*, 11, 299-307.
- Marques, A., Bruton, A., & Barney, A. (2009). Reliability of lung crackle characteristics in cystic fibrosis and bronchiectasis patients in a clinical setting. *Physiological Measurement*, 30, 903-912.
- Masterocostas, P. A., Tolia, Y. A., Theocharis, J. B., Hadjileontiadis, L. J., & Panas, S. M. (2000). An orthogonal least squares-based fuzzy filter for real-time analysis of lung sounds. *Ieee Transactions on Biomedical Engineering*, 47(9), 1165-1176.
- Murphy, R. L., Del Bono, E. A., & Davidson, F. (1989). Validation of an automatic crackle (rale) counter. *The American review of respiratory disease*, 140(4), 1017-1020.
- Musen, M., Middleton, B., & Greenes, R. (2014). Clinical Decision-Support Systems. *Biomedical Informatics*, 46(5), 643-674.
- Oliveira, A., & Marques, A. (2014). Respiratory sounds in healthy people: a systematic review. *Respir Med*, 108(4), 550-570.
- Oliveira, A., Pinho, C., Dinis, J., Oliveira, D., & Marques, A. (2013). *Automatic Wheeze Detection and Lung Function Evaluation: a preliminary study*. Paper presented at the Healthinf - 6th International Conference on Health Informatics, Barcelona.
- Ono, M., Arakawa, K., Mori, M., Sugimoto, T., & Harashima, H. (1989). Separation of Fine Crackles from Vesicular Sounds by a Nonlinear Digital-Filter. *Ieee Transactions on Biomedical Engineering*, 36(2), 286-291.
- Piirila, P. (1992). Changes in crackle characteristics during the clinical course of pneumonia. *Chest*, 102(1), 176-183.
- Quintas, J., Campos, G., & Marques, A. (2013). *Multi-algorithm respiratory crackle detection*. Paper presented at the Healthinf - 6th International Conference on Health Informatics, Barcelona.
- Rossi, M., Sovijärvi, A., Piirila, P., Vannuccini, L., & Dalmaso, F. V. J. (2000). Environmental and subject conditions and breathing manoeuvres for respiratory sound recordings. *European Respiratory Review*, 10(77), 611-615.

- Sevcik, C. (2010). A procedure to Estimate the Fractal Dimension of Waveforms. *Complexity International*, 5.
- Shen, J., Shen, W., Castan, S., & Zhang, T. X. (2002). Sum-box technique for fast linear filtering. *Signal Processing*, 82(8), 1109-1126.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation., 4304, 1015-1021.
- Sovijärvi, A., Dalmaso, F., Vanderschoot, J., Malmberg, P., Righini, G., & Stoneman, T. (2000). Definition of terms for applications of respiratory sounds. *European Respiratory Review*, 10(77), 597-610.
- Sovijärvi, A., Malmberg, L., Charbonneau, G., Vanderschoot, J., Dalmaso, F., Sacco, C., . . . Earis, J. (2000). Characteristics of breath sounds and adventitious respiratory sounds. *European Respiratory Review*, 10(77), 591–596.
- Sovijärvi, A., Vanderschoot, J., & Earis, J. (2000). *Computerized respiratory sound analysis (CORSAs): recommended standards for terms and techniques : ERS Task Force Report* (Munksgaard Ed. Vol. 10). Copenhagen.
- Sovijärvi, A. R. A., Vanderschoot, J., & Earis, J. E. (2000). Standardization of computerized respiratory sound analysis. *European Respiratory Review*, 10(77), 585.
- Taplidou, S. A., & Hadjileontiadis, L. J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. *Computers in Biology and Medicine*, 37(8), 1073-1083.
- Vannuccini, L., Earis, J., Helistö, P., Cheetham, B., Rossi, M., Sovijärvi, A., & Vanderschoot, J. (2000). Capturing and preprocessing of respiratory sounds. *European Respiratory Review*, 10(77), 616-620.
- Vannuccini, L., Rossi, M., & Pasquali, G. (1998). A new method to detect crackles in respiratory sounds. *Technology and health care*, 6(1), 75-79.
- Welsby, P., & Earis, J. (2001). Some high pitched thoughts on chest examination. *Postgraduate Medical Journal*, 77, 617-620.
- World Health Organization. (2008). The World Health Report 2008: Primary Health Care - Now more than ever. Geneva, Switzerland World Health Organization.